Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005

Unusual Evolution of Interspersed Repeat Sequences in the *Drosophila ananassae* Subgroup

Masafumi Nozawa, Masahiko Kumagai,[1] *Tadashi Aotsuka, and Koichiro Tamura*

Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan

New repeat sequences were found in the *Drosophila ananassae* genome sequence. They accounted for approximately 1.2% of the *D. ananassae* genome and were estimated to be more abundant in genomes of its closely related species belonging to the *Drosophila bipectinata* complex, whereas it was entirely absent in the *Drosophila melanogaster* genome. They were interspersed throughout euchromatic regions of the genome, usually as short tandem arrays of unit sequences, which were mostly 175–200 bp long with two distinct peaks at 180 and 189 bp in the length distribution. The nucleotide differences among unit sequences within the same array (locus) were much smaller than those between separate loci, suggesting within-locus concerted evolution. The phylogenetic tree of the repeat sequences from different loci showed that divergences between sequences from different chromosome arms occurred only at earlier stages of evolution, while those within the same chromosome arm occurred thereafter, resulting in the increase in copy number. We found RNA polymerase III promoter sequences (A box and B box), which play a critical role in retroposition of short interspersed elements. We also found conserved stem-loop structures, which are possibly associated with certain DNA rearrangements responsible for the increase in copy number within a chromosome arm. Such an atypical combination of characteristics (i.e., wide dispersal and tandem repetition) may have been generated by these different transposition mechanisms during the course of evolution.

Introduction

Recent progress in genome sequencing has revealed that repetitive sequences, e.g., transposable elements and satellite DNAs, represent a substantial fraction of a genome sequence, e.g., ~50% of human (International Human Genome Sequencing Consortium 2001), ~20% of fruit fly (Kapitonov and Jurka 2003), and ~40% of rice (Goff et al. 2002) genomes. This addresses an important issue for understanding the evolution of genomes: why and how do repetitive sequences accumulate in a genome?

Transposable elements and satellite DNAs are two major types of repetitive sequences: nearly half of the euchromatic regions in the human genome are transposable elements (International Human Genome Sequencing Consortium 2001) and heterochromatic regions harbor substantial amount of satellite DNAs (Lee et al. 1997). Transposable elements can be classified into DNA transposons and retrotransposons according to their mechanisms of transposition. DNA transposons are directly replicated from their parental genomic DNA copy, whereas retrotransposons are reverse-transcribed from their transcripts (Craig 2002 for review). Transposable elements are usually interspersed in a genome as separate copies. On the other hand, satellite and minisatellite DNAs are often found as runs of thousands or more copies of unit sequences (~100 and ~15 bp, respectively) localized in heterochromatic regions (Strachan and Read 2004, pp. 265–268, for review), whereas microsatellite DNAs (unit length is 2–5 bp) are interspersed throughout a genome, including euchromatic regions, as much shorter arrays (10–60 copies) (Litt and Luty 1989). Therefore, these different types of repetitive sequences are characterized by their lengths, distribution patterns, and copy numbers per locus.

In this study, we describe new repeat sequences found in species belonging to the *Drosophila ananassae* subgroup. They are usually occurring as tandem arrays of several units like microsatellite DNAs, while their unit length (175–200 bp) and interspersed distribution patterns are similar to those of short interspersed elements (SINEs). This unusual arrangement may hint at processes and mechanisms of repetitive-sequence increases in the number of copies per locus and in the number of loci per genome. Furthermore, their relatively recent appearance (at least after the divergence of *Drosophila melanogaster* and the *D. ananassae*) may provide us an opportunity to trace the entire evolutionary process back to their origin. Therefore, we examined the number and distribution pattern of the repeat loci in a genome, the number of units per locus, and phylogenetic relationships among the sequences from different loci, using the draft sequence of the *D. ananassae* genome (Assembly/Alignment/Annotation of 12 *Drosophila* Genomes). We also compared DNA amounts of the repeat sequences among the related species, using experimental techniques. The processes and mechanisms of the increase in the copy number of the repeat sequences are discussed.

Materials and Methods

Fly Species

Nine *Drosophila* species belonging to the *D. ananassae* subgroup were used, along with *D. melanogaster* as an outgroup species. Among these species, fly stocks of *Drosophila bipectinata*, *Drosophila malerkotliana* (*D. m. malerkotliana* and *D. m. pallens*), *Drosophila merina*,

[1] Present address: Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.

Key words: repeat sequences, satellite DNAs, transposable elements, retrotransposons, DNA transposons, *Drosophila ananassae*.

E-mail: ktamura@evolgen.biol.metro-u.ac.jp.

Mol. Biol. Evol. 23(5):981–987. 2006
doi:10.1093/molbev/msj105
Advance Access publication February 8, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Drosophila parabipectinata, *Drosophila varians*, and *D. melanogaster* (Oregon-R) were maintained at Tokyo Metropolitan University (TMU), while those of *D. ananassae* (stock number 14024-0371.0), *Drosophila pseudoananassae nigrens* (stock number 14024-0411.0), and *D. p. pseudoananassae* (stock number 14024-0421.0) were originally obtained from the National Drosophila Species Stock Center at Bowling Green State University and maintained at TMU before use.

Dot Blot Hybridization

Total DNA was extracted from 10–20 adult flies, using Boom et al.'s (1990) method with minor modifications. Next, a serial dilution series (330 pg–50 ng) of the total DNA was spotted on a sheet of nylon membrane (Amersham Biosciences, Buckinghamshire, UK) adjacent to a series of known quantities (10 pg–1 ng) of PCR amplified fragments of the repeat sequence as a quantity control. The same PCR products were also used as the probe DNA. A set of degenerate primers (5′-CGGATRMSWATTGGVVAAGATA-3′ and 5′-SCCAATWNKYATCCRATTCT-3′) were used for the PCR to amplify a variety of sequence variants. The hybridization, washing, and detection were conducted according to the protocol of Alkphos Direct Labelling and Detection System with CDP-star (Amersham Biosciences). The signal strength was measured by using a ChemiDoc image analyzer with Quantity-One software (Bio Rad, Hercules, Calif.). These procedures were replicated three times for each species and the average signal strength was calculated.

Fluorescence In Situ Hybridization

Salivary glands were excised from third instar larvae of *D. parabipectinata* in a drop of 40% acetic acid, soaked in fixative (16% lactic and 50% acetic acids) for 20 min, and spread on a glass slide by squashing. These preparations were then hybridized with probe DNA at 55°C for 4 h. For the hybridization probe, we used the sequence of a repeat locus located between two *siren* loci of *D. parabipectinata* (nucleotide position 2922–3891 of accession number AB194414), which was cloned with pUC118 as the vector and labeled with DIG-Nick Translation Mix (Roche Diagnostics, Indianapolis, Ind.). After the hybridization, the probe DNA was detected using Anti-Digoxigenin-Fluorescein Fab fragments (Roche Diagnostics) and visualized with an Olympus BX60 fluorescence microscope.

Data Analyses of the *D. ananassae* Genome Sequence

The number of repeat units per locus and the distribution of repeat loci in a genome were examined by using the draft sequence of the *D. ananassae* genome (2004.12.6 release) downloaded from: http://rana.lbl.gov/drosophila/assemblies/dana_agencourt_arachne_06dec04.tar.gz. To take into account the sequence variation, we performed a two-round BlastN (Altschul et al. 1990) search procedure. The first round search was aimed to find a representative sequence for the query in the second round search. The repeat sequence between two *siren* loci of *D. parabipectinata* (nucleotide position 2921–4152, accession no. AB194414) was used as the query. As a result, we obtained 2,881 target sequences at the expected value (E-value) $\leq 10^{-5}$. From these sequences, five were arbitrarily chosen, each of which was used as a query for the second round BlastN search with E-value $\leq 10^{-5}$. In the five sets of sequences obtained, the largest set (including 3,165 sequences from 2,340 separate contigs) was employed as the initial data set because a representative sequence is expected to hit more target sequences. This data set covered 94% of the repeat sequences found by all five queries. However, this initial data set likely contained overlaps because they came from many separate contigs.

To extract only unique loci from the initial data set, we mapped each repeat sequence on the complete *D. melanogaster* genome sequence (release 4.0) using both 5′ and 3′ flanking sequences (1 kb) as the queries for the BlastN search with E-value $\leq 10^{-10}$. The correspondence between *D. ananassae* and *D. melanogaster* chromosome arms is well established: i.e., X, 2L, 2R, 3L, 3R, and 4 of *D. ananassae* correspond to X, 3R, 3L, 2R, 2L, and 4 of *D. melanogaster*, respectively (Ashburner 1989, p. 23; M. Matsuda, personal communication). So far, no between-chromosome-arm rearrangement has been found between *D. ananassae* and *D. melanogaster* by means of careful inspections of the banding patterns of polytene chromosomes except for rRNA gene clusters (Kikkawa 1938; Roy et al. 2005; M. Matsuda, personal communication).

Because the first and last sites of repeat sequences are not always the same among different loci, it is relatively complex to determine the unit sequence. For instance, when there are two types of sequences, e.g., ATATAT… and TATATA…, the unit sequence of the dinucleotide repeat may be either AT or TA. For satellite DNAs, the first site of unit sequences has often been defined arbitrarily by using a certain landmark sequence like a restriction site (e.g., *Kpn*I family defined by Kazama et al. 2003). In this study, we used a majority consensus rule to define the unit sequence, i.e., the first site was determined as the site where the repeat sequences start most frequently in the *D. ananassae* genome.

Phylogenetic Analyses of the Repeat Sequences

A phylogenetic tree of the *D. ananassae* repeat sequences was constructed by the minimum-evolution method (Rzhetsky and Nei 1992). The proportion of differences (p-distance) was used as the evolutionary distance with the pairwise deletion option after sequences were aligned by ClustalW (Thompson, Higgins, and Gibson 1994). MEGA3 software (Kumar, Tamura, and Nei 2004) was used to carry out these procedures. The rate of increase in the copy number was computed by dividing the estimated copy numbers by the divergence times among the species of the *D. ananassae* subgroup. The divergence times were estimated from the *Adh* tree (Nozawa, Aotsuka, and Tamura 2005) by using the linearized-tree method (Takezaki, Rzhetsky, and Nei 1995) under the assumption that *D. ananassae* and *D. melanogaster* diverged at 44.2 MYA (Tamura, Subramanian, and Kumar 2004).

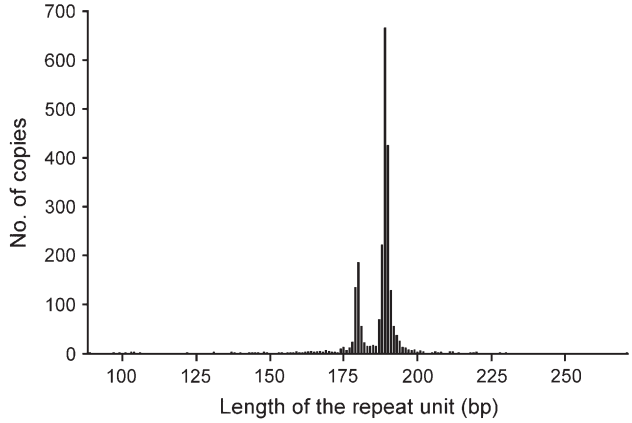

FIG. 1.—Distribution of the unit lengths of the repeat sequences.

Results

The Number and Distribution of Repeat Loci in the *D. ananassae* Genome

Using the two-round BlastN searches, we found 3,165 repeat loci in the draft sequence of *D. ananassae* genome. Because these repeat loci were scattered across 2,340 separate contigs, some of them are likely overlapping on the same loci. To remove the redundancy and extract only unique loci, we mapped each locus on the *D. melanogaster* complete genome sequence using its 5′ and 3′ flanking sequences (1 kb) as queries. Both flanking sequences were available for only 1,364 repeat sequences, and 831 of them were consistently mapped on particular chromosomal locations by both of their flanking sequences. Among these 831 sequences, 16 overlapped others; hence, 815 unique loci were ultimately identified.

The unit sequence of the repeat was defined as the sequence between the sites at which the repeat sequences start most frequently in the *D. ananassae* genome (supplementary fig. 1). The length of the repeat unit mostly ranged from 175 to 200 bp (89 at minimum and 271 at maximum) varied by short duplications and deletions (fig. 1). However, there are two distinct peaks at 180 and 189 bp in the distribution, suggesting all repeat sequences can be classified into two types: short (S) and long (L). The distribution of the number of units per locus (fig. 2) shows that there are two or more

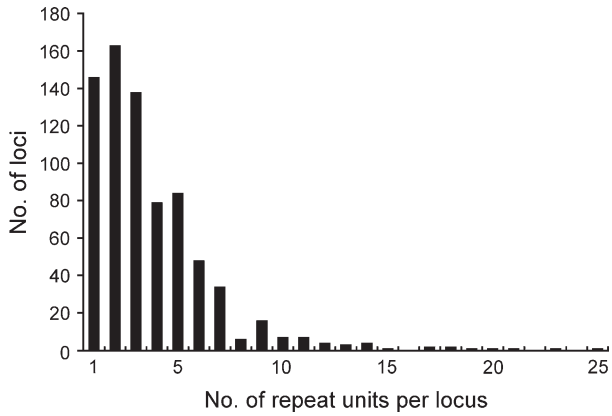

FIG. 2.—Distribution of the number of repeat units per locus.

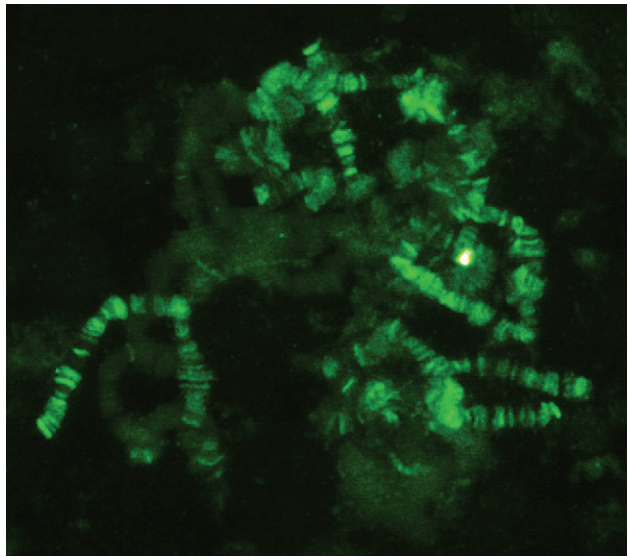

FIG. 3.—In situ hybridization of the repeat sequences to polytene chromosomes of *Drosophila parabipectinata*. The wide distribution of the repeat sequences is indicated by the green stain.

units in 603 out of the 815 loci. In the remaining 212 loci, there are only single (146 loci) or partial (66 loci) units. Given that there are 2,823 repeat units in the 815 loci, the number of repeat units per locus is 3.5 on average.

Repeat loci are distributed on all autosomal arms almost uniformly, but occur less frequently on the X chromosome; there are 4.4, 7.7, 7.5, 7.5, and 7.5 loci per 1 Mb genome sequence for the X chromosome and 2L, 2R, 3L, and 3R chromosome arms, respectively. This broadly interspersed distribution of loci throughout euchromatic regions is further illustrated by the green stain in the in situ hybridization to polytene chromosomes of *D. parabipectinata* (fig. 3).

Phylogenetic Relationships Among Repeat Sequences

The number of nucleotide substitutions per site between repeat sequences within a locus is 0.07 on average, which is substantially smaller than that between loci (0.34), suggesting that repeat sequences in the same locus have evolved in a concerted manner. Therefore, for examining the phylogenetic relationships among loci, we used only a single unit sequence from each locus. In the phylogenetic tree obtained, the repeat sequences from the same chromosome arm are not monophyletically clustered (fig. 4). However, there is a clear trend in that sequences from the X chromosome and 2R chromosome arm are localized in clusters A and B, respectively. Actually, chromosomal distributions of the repeat sequences show a significant heterogeneity among clusters A, B, and C (table 1). The close relationships among the sequences in cluster A are also supported by the fact that all but 2, out of 41, sequences are the S type. It is noteworthy that common ancestors of the repeat sequences currently located in different chromosome arms (displayed by thick black lines in fig. 4) are all old, whereas the recent divergences are confined between sequences located within the same chromosome arm. This

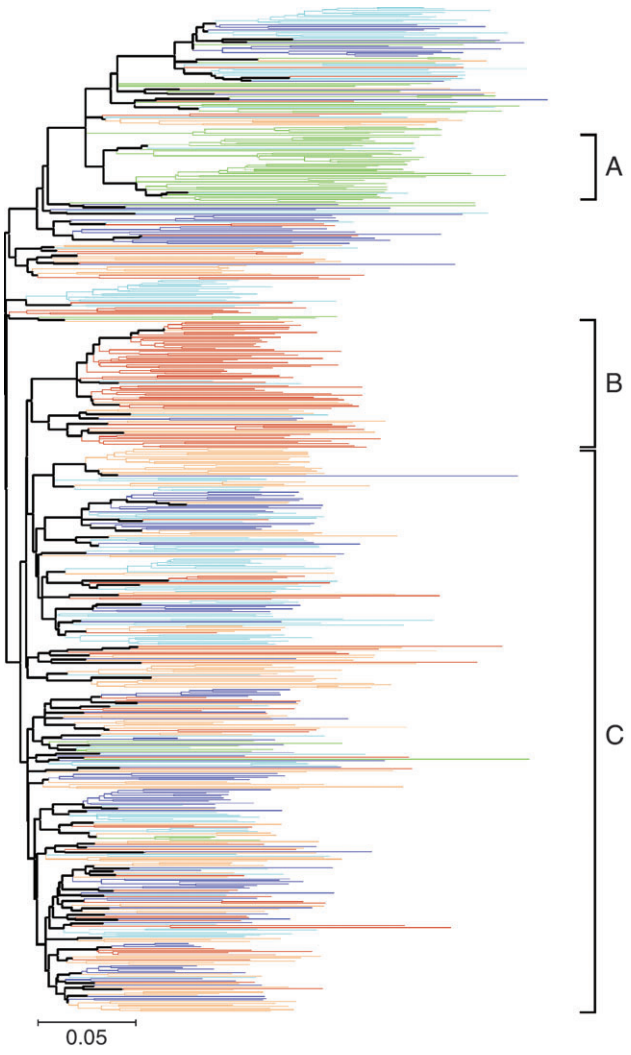

FIG. 4.—Minimum-evolution tree of the repeat sequences. The first intact repeat unit from each locus was used. Lineages leading to the repeat sequences on the X chromosome, 2L, 2R, 3L, and 3R chromosome arms are drawn by green, pink, red, light blue, and blue lines, respectively, whereas lineages leading to the repeat sequences on more than one chromosome arms are drawn by thick black lines. See text for clusters A, B, and C.

suggests that the interspersed chromosomal distribution of loci was established during earlier evolutionary processes and the later divergences within each chromosome arm have contributed to the increase in the number of copies.

Table 1
Chromosomal Distribution of Repeat Loci in the *Drosophila ananassae* Genome among Clusters A, B, and C of the Phylogenetic Tree

	The Numbers of the Repeat Loci Among Chromosome Arms				
Clusters[a]	X	2L	2R	3L	3R
A	39	0	0	2	0
B	0	11	62	3	2
C	9	120	47	82	91

NOTE.—The distributions of repeat loci are significantly different between all three pairs of clusters ($P \ll 0.0001$ in χ^2 test).

[a] The clusters A, B, and C are defined in the phylogenetic tree in figure 4.

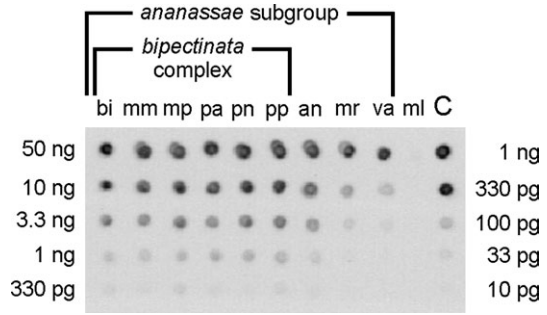

FIG. 5.—Dot blot hybridization of the repeat sequences from the species of the *Drosophila ananassae* subgroup. The amount of the total DNA from each species is listed in the left margin, whereas the amount of the control DNA (PCR products of the repeat sequences) is indicated in the right margin. bi, *Drosophila bipectinata*; mm, *Drosophila m. malerkotliana*; mp, *Drosophila m. pallens*; pb, *Drosophila parabipectinata*; pn, *Drosophila p. nigrens*; pp, *Drosophila p. pseudoananassae*; an, *D. ananassae*; mr, *Drosophila merina*; va, *Drosophila varians*; ml, *Drosophila melanogaster*; C, control.

Evolution of the Repeat Sequences in the *D. ananassae* Subgroup

The total amounts of repeat sequence DNA in a genome were measured by dot blot hybridization and compared among the species of the *D. ananassae* subgroup and *D. melanogaster* (fig. 5). To take into account the sequence variation, we used a heterogeneous probe, which was amplified from the total DNA of *D. ananassae* by PCR with a degenerate primer set (see *Materials and Methods* for details). Positive signals were detected in all the *D. ananassae* subgroup species examined. The amount of repeat sequence DNA was estimated to be 0.5%–1.5% in *D. ananassae*, *D. merina*, and *D. varians*. This is consistent with the value (1.2%) obtained by the BlastN search of the *D. ananassae* genome sequence. The amount was larger (2%–3%) in species belonging to the *D. bipectinata* complex. However, for *D. melanogaster*, it was below the limit of experimental detection (0.02% at best). This is consistent with the observation that the BlastN search did not hit the repeat sequence in the complete *D. melanogaster* genome sequence (release 4.0).

Assuming an average genome size for the *D. ananassae* subgroup of 220 Mb (0.57 pg DNA per diploid cell

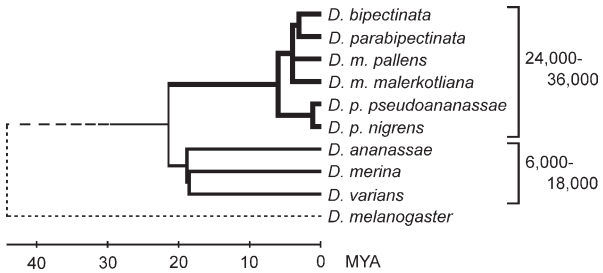

FIG. 6.—A linearized *Adh* gene tree showing the phylogenetic relationships of the species in the *Drosophila ananassae* subgroup with *Drosophila melanogaster* as an outgroup. The dotted lines indicate the absence of the repeat sequence, while the thickness of the solid lines corresponds to the total copy number of repeat sequences in a genome described in the right margin.

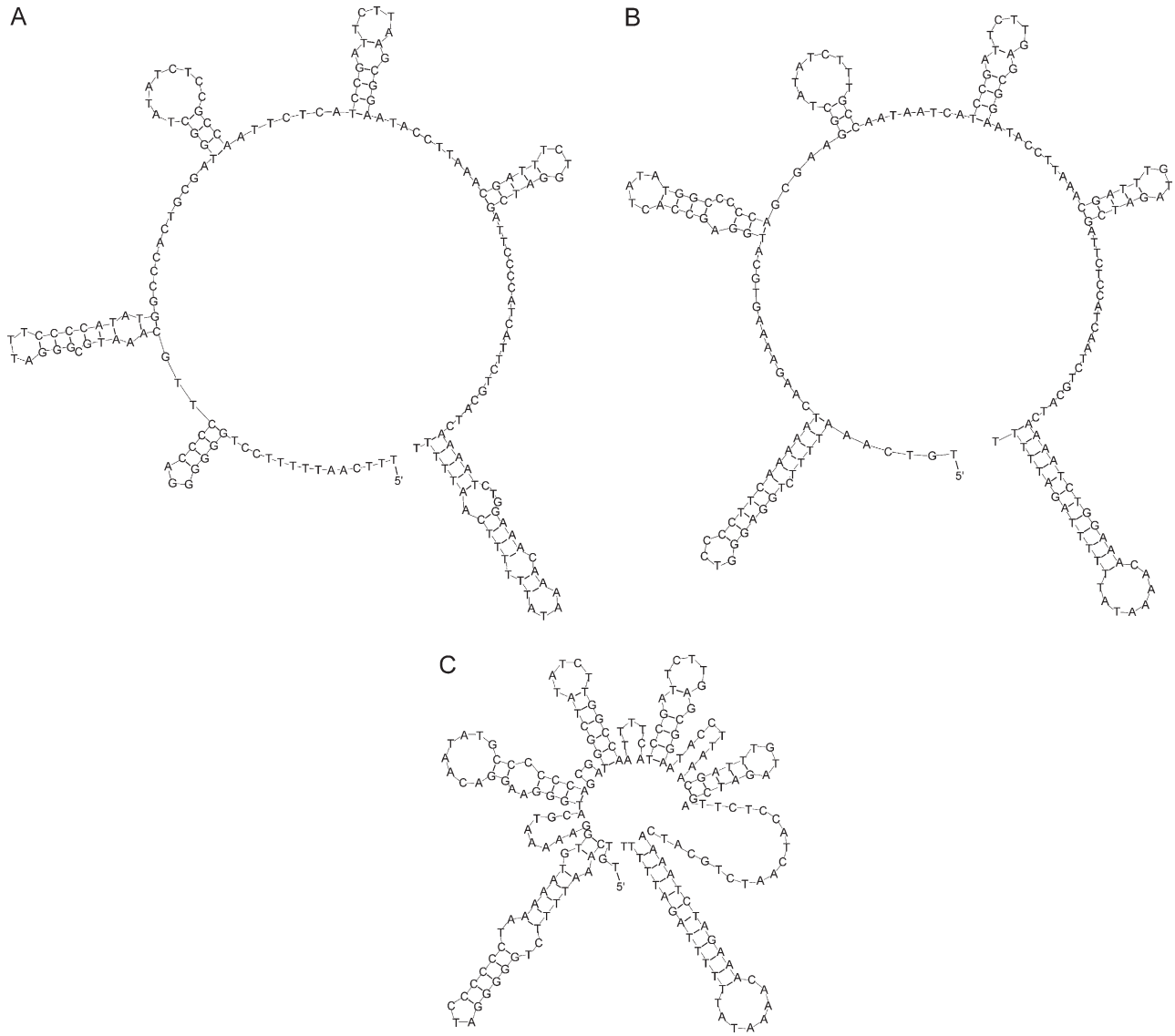

FIG. 7.—Secondary structure of the repeat sequences. Optimal folding structure was estimated by using MFOLD program (Zuker, Mathews, and Turner 1999) from the consensus unit sequences for clusters A, B, and C shown in figure 4.

according to Nardon et al. 2005) and a 185-bp average length for the repeat units, we estimated the total copy numbers of the unit sequences in a genome from the repeat sequence DNA amounts obtained. Figure 6 shows the *Adh* gene tree of the *D. ananassae* subgroup species (Nozawa, Aotsuka, and Tamura 2005) and their estimated total copy numbers of the repeat sequences in the right margin. Because the BlastN search identifies the repeat sequence in neither *D. melanogaster* nor *Drosophila pseudoobscura*, it is thought to be absent in the ancestral species of the *D. melanogaster* group (broken line), but arose and increased in copy number during the evolution of the *D. ananassae* subgroup after divergence from *D. melanogaster* (solid line). At present, the copy numbers have become 6,000–18,000 in *D. ananassae*, *D. merina*, and *D. varians* (thick lines) and 24,000–36,000 in the *D. bipectinata* complex (thicker lines). Therefore, the average rate of copy number increase can be estimated to be 24,000–36,000/44.2 $\approx$ 540–820 copies/million year for the evolution of the *D. bipectinata* complex.

Discussion

In this study, the evolution of the repeat sequences found in the *D. ananassae* genome was investigated. They are widely interspersed in euchromatic regions, as are many retrotransposons (Kidwell 2005). The length of their unit sequences is particularly similar to that of SINEs (Graur and Li 2000, pp. 343–349). If the repeat sequences are a type of SINEs, they should contain the characteristic sequences of the so-called A box and B box, which function together as a promoter of RNA polymerase III (Murphy and Baralle 1983). Indeed, 3 out of 2,823 sequences have both A and B boxes, while 17 have only A box and 1,020 have only B box. Another characteristic of retrotransposons is that the genomic location of a newly generated copy has nothing

to do with the location of the parental copy; this was demonstrated by the close phylogenetic relationships among sequences from distant chromosomal positions (Cardazzo et al. 2003). In our phylogenetic tree (fig. 4), close relationships among repeat sequences from different chromosome arms are exclusively observed in old lineages, while only within-chromosome-arm transpositions have occurred recently. This implies that the repeat sequences were distributed by a certain reverse transcription activity over various chromosomal locations in the earlier stage of the evolution, but the activity was reduced thereafter and currently no longer in effect.

However, there is a feature in the repeat sequences that distinguishes them from SINEs, i.e., they frequently occur as tandem arrays of two or more copies (fig. 2). This situation is quite rare for euchromatic SINEs. For instance, ~97% of *Alu* elements, which are the most abundant SINEs in the human genome, occur as single copies in euchromatic regions (El-Sawy and Deininger 2005). Although some transposable elements form tandem arrays, they are usually much longer and localized in heterochromatic regions (Wong and Choo 2004) as in satellite DNAs (Strachan and Read 2004, pp. 265–268). The recent within-chromosome-arm divergences of the repeat sequences shown in the phylogenetic tree (fig. 4) are consistent with the evolution of satellite DNAs and DNA transposons. For instance, there are chromosome-specific subfamilies in PIM357 (Pons, Juans, and Petitpierre 2002) and α (Choo et al. 1991) satellite DNAs. A strong tendency for adjacent transpositions in the same chromosome is known in various kinds of DNA transposons, e.g., *Ac* (Machida et al. 1997), *Ds* (Bancroft and Dean 1993), *Hermes* (Guimond et al. 2003), and *P* (Tower et al. 1993) elements. In this context, Guimond et al. (2003) argued that "local hopping" is a common characteristic of DNA transposons. Because the within-chromosome-arm divergences are prominent in clusters A and B (fig. 4), we compared the consensus sequences of these clusters with each other and with that of cluster C to identify the sequence responsible for the within-chromosome-arm transpositions. The sequence similarity between the clusters A and B is 81%, whereas those between A and C and between B and C are 82% and 90%, respectively. The higher similarity between the clusters B and C is consistent with the similarity in the unit length; most of the sequences in cluster A are of the S type, whereas most of the sequences in clusters B and C are of the L type. Therefore, there is no clear commonality found in the primary structure of the sequences between clusters A and B as compared with cluster C. However, in terms of the optimal folding structure deduced by MFOLD program (Zuker, Mathews, and Turner 1999), the consensus sequences of clusters A and B show a higher similarity to each other than to those from the consensus sequence of cluster C, although the major stem-loop structures are well conserved in all the three sequences (fig. 7). This situation is similar with results of a comparative study of satellite DNAs found in the flour beetle (the genus *Tribolium*), which indicates a common secondary structure despite a low sequence similarity (Ugarković and Plohl 2002). Kato, Matsunaga, and Shimizu (1998) suggested that such secondary structures may play a role in the increase of satellite DNAs. In addition, secondary structures are likely associated with DNA transpositions (Izsvák et al. 1999). Therefore, the observed secondary structure is possibly responsible for the recent increase within a chromosome arm.

In conclusion, we propose a plausible scenario to explain the occurrence and distribution of the repeat sequences in the *D. ananassae* subgroup genomes. During the course of evolution of the *D. ananassae* subgroup, the mechanism of the repeat sequence increase has itself changed. Retroposition likely caused the wide dispersal throughout euchromatic regions during the early stages of evolution, and subsequently, DNA rearrangements contributed to the increase in copy numbers within each chromosome arm. This unusual evolutionary history seems to have resulted in an atypical pattern of chromosomal distribution, a state of tandem repetition within a locus, and phylogenetic relationships among sequences described in this paper.

Supplementary Material

Supplementary figure 1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Muneo Matsuda and Yoshihiko Tomimura for their invaluable advices during the course of this work. We also thank Akihiko Koga for information about transposition. Additional thanks to G. Curt Fiedler, who proofread the final manuscript and gave us helpful comments. This work was supported by grants from the Japan Society for the Promotion of Science to T.A. (12375002 and 15255006) and K.T.

Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

Ashburner, M. 1989. *Drosophila*: a laboratory handbook. Cold Spring Harbor Laboratory Press, New York.

Bancroft, I., and C. Dean. 1993. Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana* Genetics **134**:1221–1229.

Boom, R., C. J. A. Sol, M. M. M. Salimans, C. L. Jansen, P. M. E. Wertheim-van Dillen, and J. Van Der Noordaa. 1990. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. **28**:495–503.

Cardazzo, B., L. Bargelloni, L. Toffolatti, P. Rimessi, A. Ferlini, and T. Patarnello. 2003. Tempo and mode of evolution of a primate-specific retrotransposon belonging to the LINE 1 family. J. Mol. Evol. **57**:S268–S276.

Choo, K. H., B. Vissel, A. Nagy, E. Earle, and P. Kalitsis. 1991. A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. **19**:1179–1182.

Craig, N. L. 2002. Mobile DNA: an introduction. Pp. 3–11 *in* N. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz, eds. Mobile DNA II. ASM Press, Washington, D.C.

El-Sawy, M., and P. Deininger. 2005. Tandem insertions of Alu elements. Cytogenet. Genome Res. **108**:58–62.

Goff, S. A., D. Ricke, T.-H. Lan et al. (55 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296**:92–100.

Graur, D., and W.-H. Li. 2000. Fundamentals of molecular evolution. 2nd edition. Sinauer Associates Inc., Sunderland, Mass.

Guimond, N., D. K. Bideshi, A. C. Pinkerton, P. W. Atkinson, and D. A. O'Brochta. 2003. Patterns of *Hermes* transposition in *Drosophila melanogaster*. Mol. Genet. Genomics **268**:779–790.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

Izsvák, Z., Z. Ivics, N. Shimoda, D. Mohn, H. Okamoto, and P. B. Hackett. 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol. **48**:13–21.

Kapitonov, V. V., and J. Jurka. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. USA **100**:6569–6574.

Kato, M., K. Matsunaga, and N. Shimizu. 1998. A novel unusual DNA structure formed in an inverted repeat sequence. Biochem. Biophys. Res. Commun. **246**:532–534.

Kazama, Y., R. Sugiyama, S. Matsunaga, F. Shibata, W. Uchida, M. Hizume, and S. Kawano. 2003. Organization of the *Kpn*I family of chromosomal distal-end satellite DNAs in *Silene latifolia*. J. Plant Res. **116**:317–326.

Kidwell, M. G. 2005. Transposable elements. Pp. 165–221 *in* T. R. Gregory, ed. The evolution of the genome. Elsevier Academic Press, London.

Kikkawa, H. 1938. Studies on the genetics and cytology of *Drosophila ananassae*. Genetica **20**:458–516.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5**:150–163.

Lee, C., R. Wevrick, R. B. Fisher, M. A. Ferguson-Smith, and C. C. Lin. 1997. Human centromeric DNAs. Hum. Genet. **100**:291–304.

Litt, M., and J. A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Genet. **44**:397–401.

Machida, C., H. Onouchi, J. Koizumi, S. Hamada, E. Semiarti, S. Torikai, and Y. Machida. 1997. Characterization of the transposition pattern of the *Ac* element in *Arabidopsis thaliana* using endonuclease I-*Sce*I. Proc. Natl. Acad. Sci. USA **94**:8675–8680.

Murphy, M. H., and F. E. Baralle. 1983. Directed semisynthetic point mutational analysis of an RNA polymerase III promoter. Nucleic Acids Res. **11**:7695–7700.

Nardon, C., G. Deceliere, C. Lœvenbruck, M. Weiss, C. Vieira, and C. Biémont. 2005. Is genome size influenced by colonization of new environments in dipteran species? Mol. Ecol. **14**:869–878.

Nozawa, M., T. Aotsuka, and K. Tamura. 2005. A novel chimeric gene, *siren*, with retroposed promoter sequence in the *Drosophila bipectinata* complex. Genetics **171**:1719–1727.

Pons, J., C. Juans, and E. Petitpierre. 2002. Higher-order organization and compartmentalization of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. Chromosome. Res. **10**:597–606.

Roy, V., L. Monti-Dedieu, N. Chaminade, S. Siljak-Yakovlev, S. Aulard, and F. Lemeunier. 2005. Evolution of the chromosomal location of rDNA genes in two *Drosophila* species subgroups: *ananassae* and *melanogaster*. Heredity **94**: 388–395.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**:945–967.

Strachan, T., and A. P. Read. 2004. Human molecular genetics. 3rd edition. Garland Publishing, New York.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12**:823–833.

Tamura, K., S. Subramanian, and S. Kumar. 2004. Temporal pattern of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol. Biol. Evol. **21**:36–44.

Thompson, J., D. G. Higgins, and T. I. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific fap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

Tower, J., G. H. Karpen, N. Craig, and A. C. Spradling. 1993. Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. Genetics **133**:347–359.

Ugarković, Đ., and M. Plohl. 2002. Variation in satellite DNA profiles—causes and effects. EMBO J. **21**:5955–5959.

Wong, L. H., and K. H. A. Choo. 2004. Evolutionary dynamics of transposable elements at the centromere. Trends Genet. **20**:611–616.

Zuker, M., D. H. Mathews, and D. H. Turner. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. Pp. 11–33 *in* J. Barciszewski and B.F.C. Clark, eds. RNA biochemistry and biotechnology. Kluwer Academic Publishers, Boston.

Marta Wayne, Associate Editor

Accepted February 1, 2006